EP32848(7)

(12) UK Patent Application (19) GB (11) 2 365 188 (13) A

(43) Date of A Publication 13.02.2002

(21) Application No 0017887.1

(22) Date of Filing 20.07.2000

(71) Applicant(s)
Canon Kabushiki Kaisha
(Incorporated in Japan)
3-30-2 Shimomaruko, Ohta-Ku, Tokyo 146, Japan

(72) Inventor(s)
Paul Michael Tapper

(74) Agent and/or Address for Service
Beresford & Co
2-5 Warwick Court, High Holborn, LONDON,
WC1R 5DH, United Kingdom

(51) INT CL^7
G10L 15/22 // G10L 101:18

(52) UK CL (Edition T)
G4R RHA R1F
U1S S1820 S1831 S2204 S2210 S2215 S2322

(56) Documents Cited
GB 2098773 A US 5781887 A

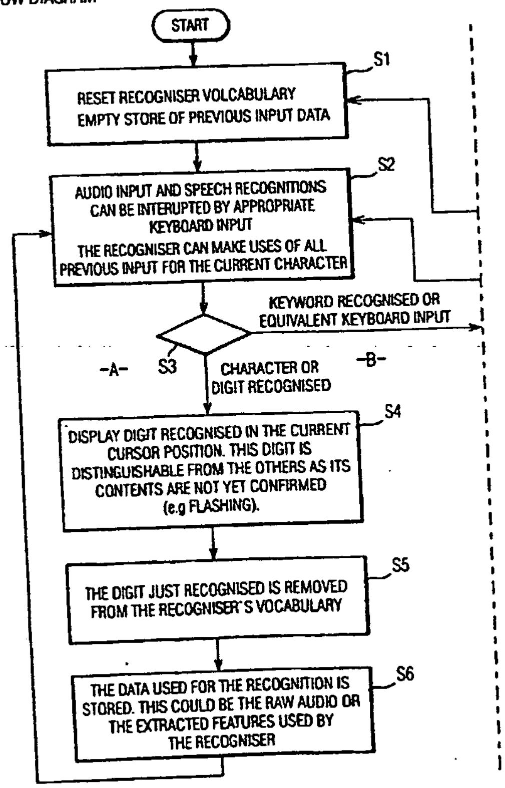
(58) Field of Search
UK CL (Edition S) G4R RHA
INT CL^7 G10L 15/00 15/22
Online:WPI, EPODOC, JAPIO

(54) Abstract Title
Entering characters by speech recognition

(57) A voice-activated data input system receives an input speech signal representing a character and compares a processed speech signal with a stored library of characters in order to allot a character from the library to the input speech signal. The allotted character is displayed, 54. If the character is incorrect the user repeats it and a different character is displayed. If the character is correct the user confirms it by saying a keyword, or by a keyboard input.

FLOW DIAGRAM

FIG.3A


START
S1
RESET RECOGNISER VOLCABULARY
EMPTY STORE OF PREVIOUS INPUT DATA
S2
AUDIO INPUT AND SPEECH RECOGNITIONS CAN BE INTERUPTED BY APPROPRIATE KEYBOARD INPUT
THE RECOGNISER CAN MAKE USES OF ALL PREVIOUS INPUT FOR THE CURRENT CHARACTER
KEYWORD RECOGNISED OR EQUIVALENT KEYBOARD INPUT
-A-
S3
CHARACTER OR DIGIT RECOGNISED
-B-
S4
DISPLAY DIGIT RECOGNISED IN THE CURRENT CURSOR POSITION. THIS DIGIT IS DISTINGUISHABLE FROM THE OTHERS AS ITS CONTENTS ARE NOT YET CONFIRMED (e.g FLASHING).
S5
THE DIGIT JUST RECOGNISED IS REMOVED FROM THE RECOGNISER'S VOCABULARY
S6
THE DATA USED FOR THE RECOGNITION IS STORED. THIS COULD BE THE RAW AUDIO OR THE EXTRACTED FEATURES USED BY THE RECOGNISER

GB 2 365 188 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

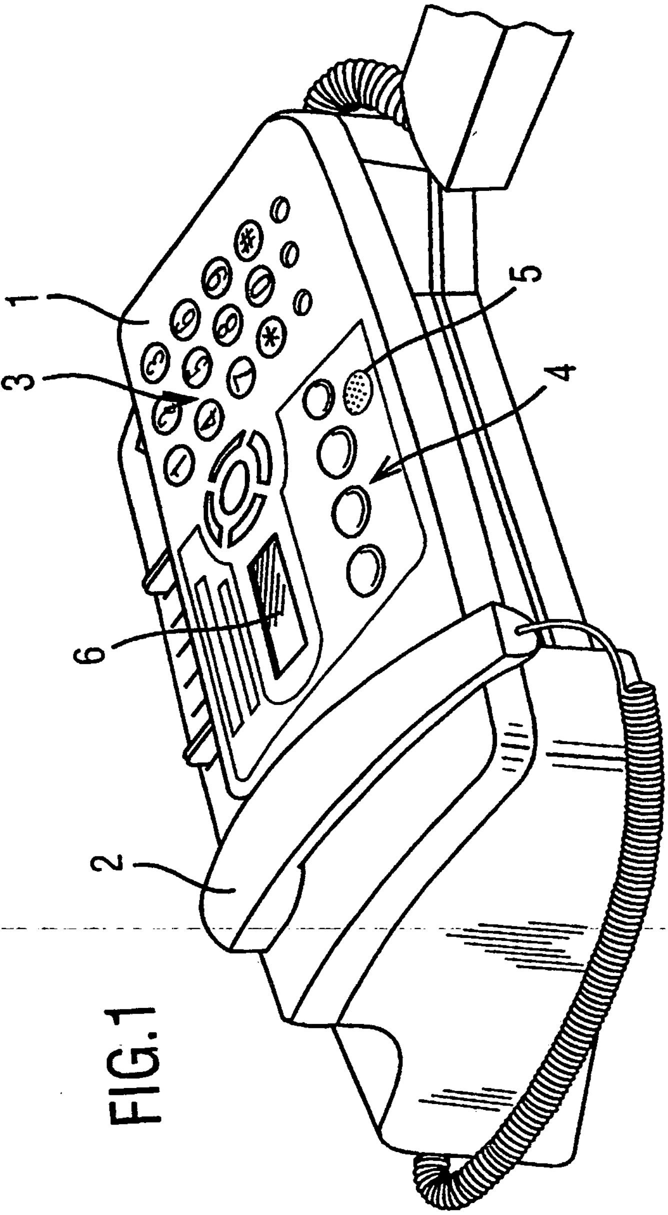


FIG.1

BLOCK DIAGRAM

FIG.2

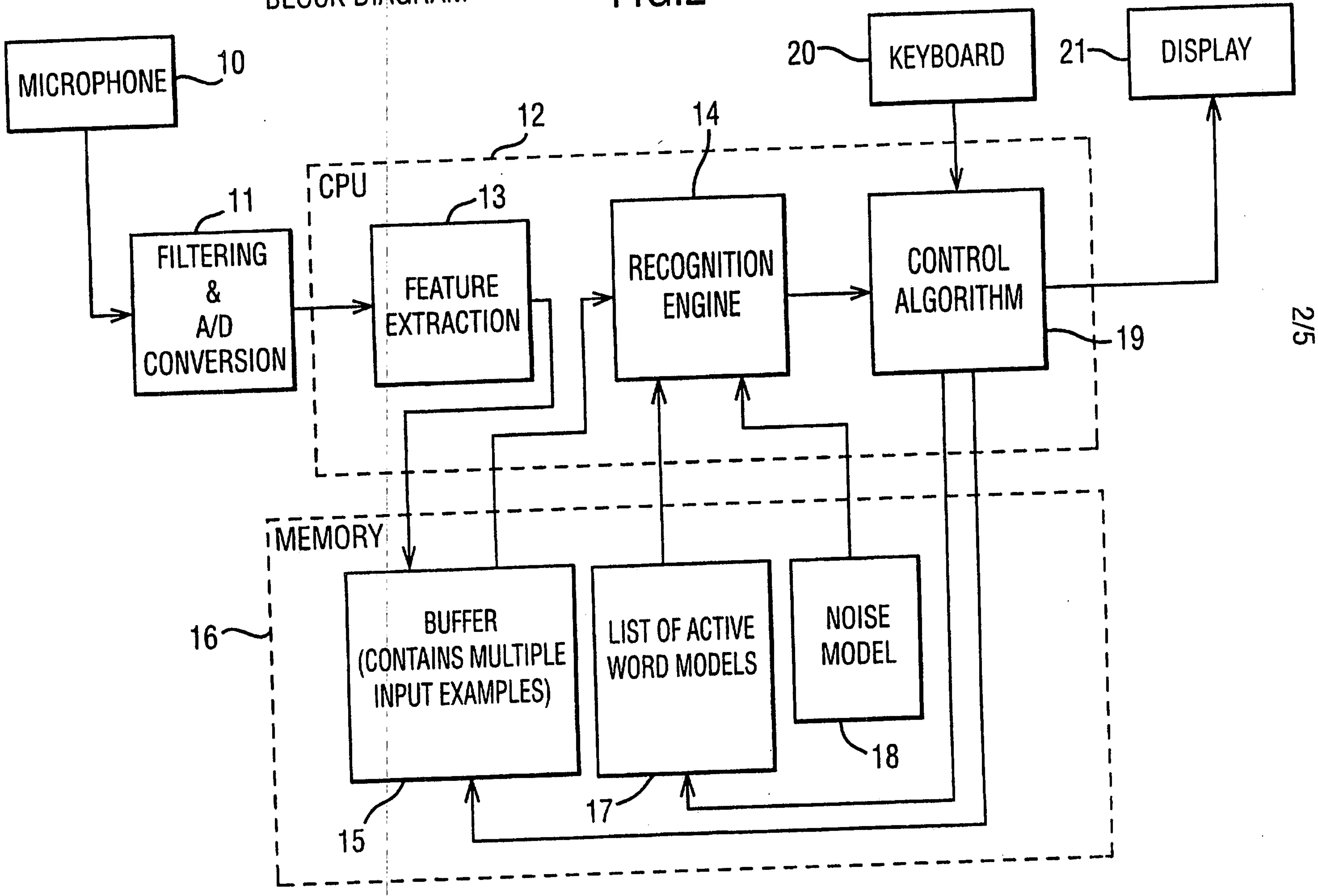


FIG.3

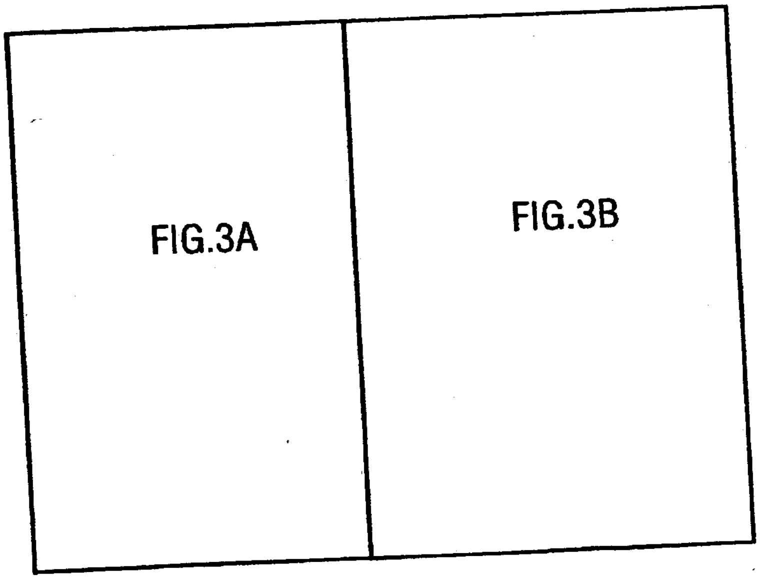
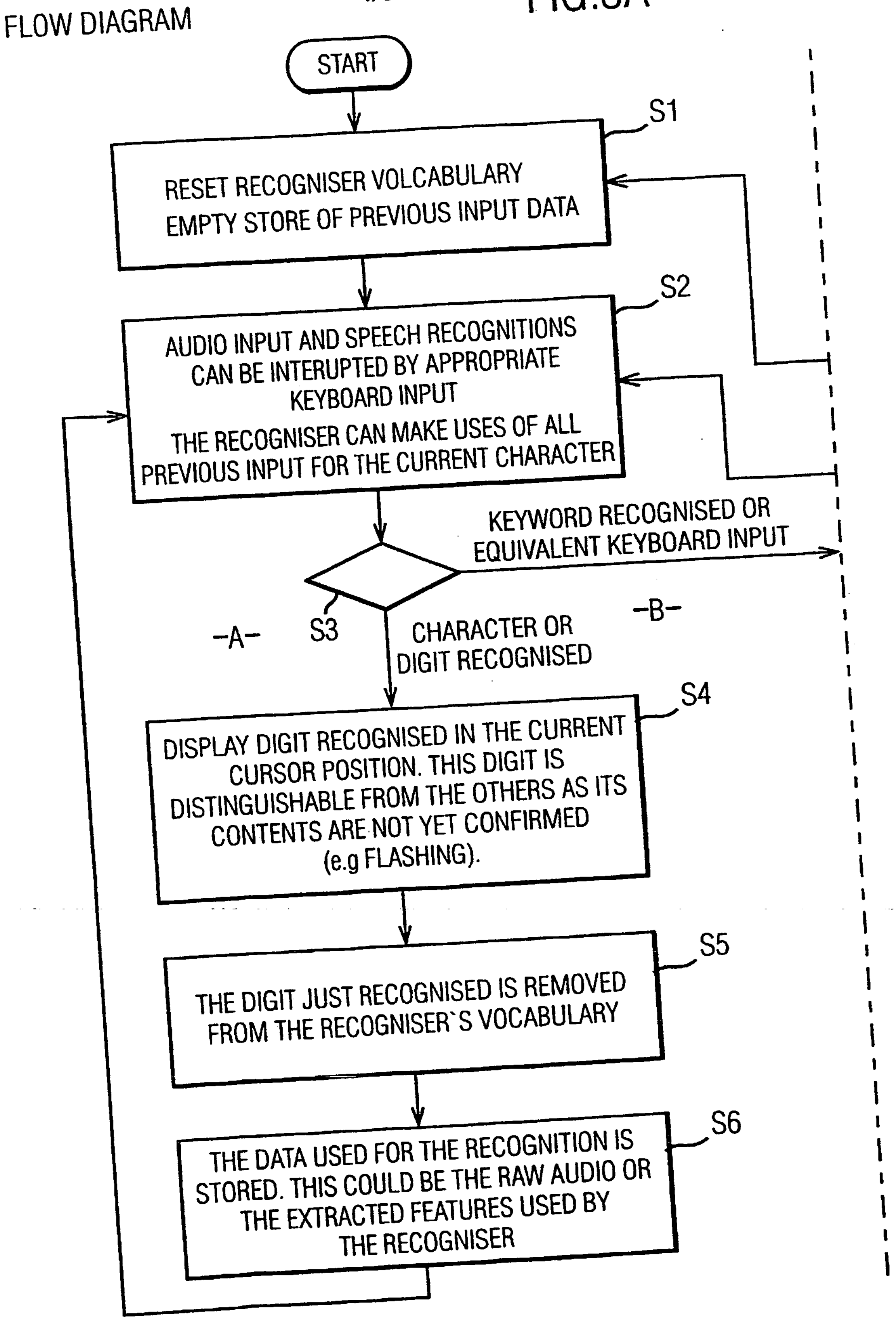
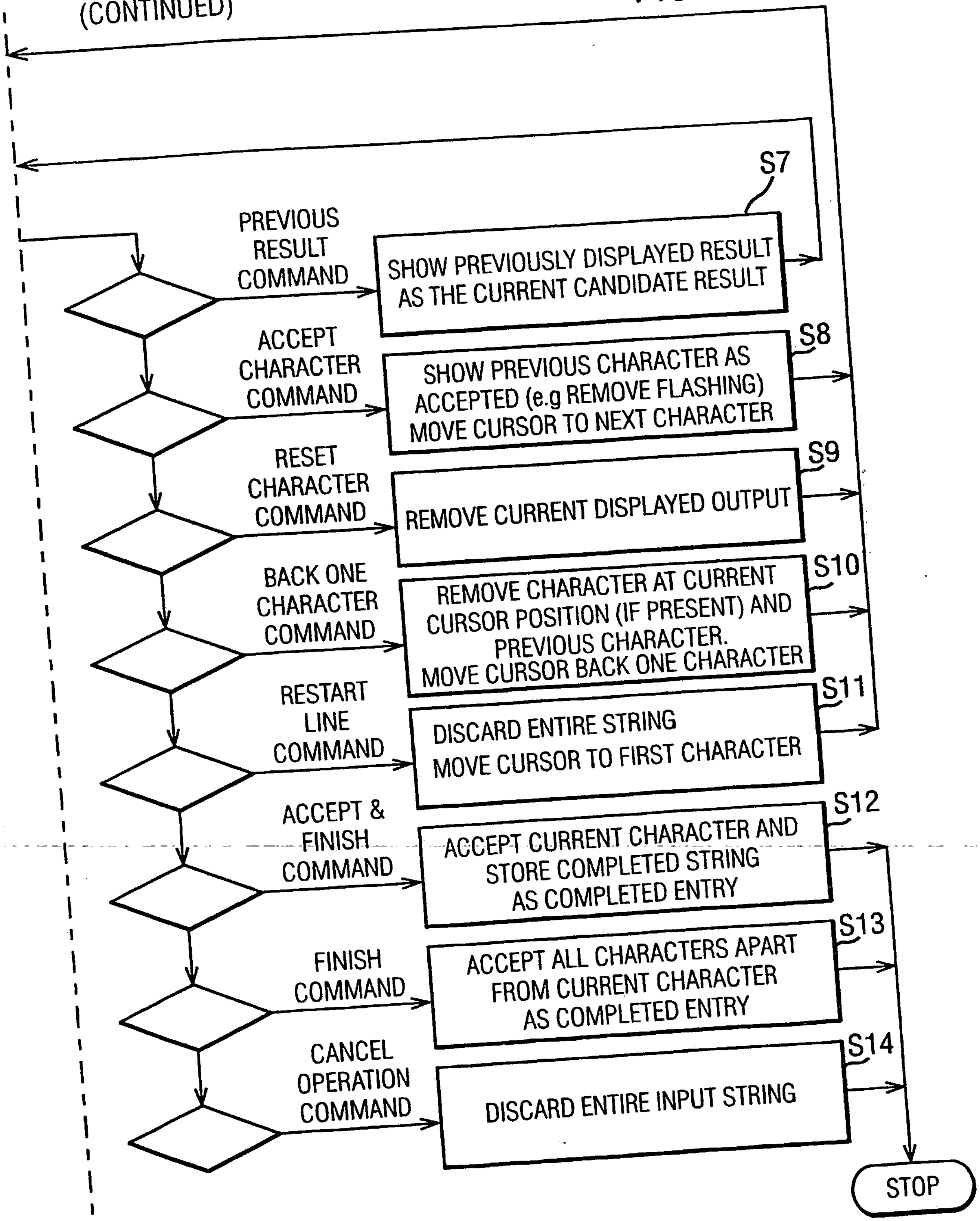


FIG.3A


FLOW DIAGRAM
START
S1
RESET RECOGNISER VOLCABULARY
EMPTY STORE OF PREVIOUS INPUT DATA
S2
AUDIO INPUT AND SPEECH RECOGNITIONS CAN BE INTERUPTED BY APPROPRIATE KEYBOARD INPUT
THE RECOGNISER CAN MAKE USES OF ALL PREVIOUS INPUT FOR THE CURRENT CHARACTER
KEYWORD RECOGNISED OR EQUIVALENT KEYBOARD INPUT
-A-
S3
CHARACTER OR DIGIT RECOGNISED
-B-
S4
DISPLAY DIGIT RECOGNISED IN THE CURRENT CURSOR POSITION. THIS DIGIT IS DISTINGUISHABLE FROM THE OTHERS AS ITS CONTENTS ARE NOT YET CONFIRMED (e.g FLASHING).
S5
THE DIGIT JUST RECOGNISED IS REMOVED FROM THE RECOGNISER`S VOCABULARY
S6
THE DATA USED FOR THE RECOGNITION IS STORED. THIS COULD BE THE RAW AUDIO OR THE EXTRACTED FEATURES USED BY THE RECOGNISER

FLOW DIAGRAM
(CONTINUED)

FIG.3B


PREVIOUS RESULT COMMAND
S7
SHOW PREVIOUSLY DISPLAYED RESULT AS THE CURRENT CANDIDATE RESULT
ACCEPT CHARACTER COMMAND
S8
SHOW PREVIOUS CHARACTER AS ACCEPTED (e.g REMOVE FLASHING) MOVE CURSOR TO NEXT CHARACTER
RESET CHARACTER COMMAND
S9
REMOVE CURRENT DISPLAYED OUTPUT
BACK ONE CHARACTER COMMAND
S10
REMOVE CHARACTER AT CURRENT CURSOR POSITION (IF PRESENT) AND PREVIOUS CHARACTER. MOVE CURSOR BACK ONE CHARACTER
RESTART LINE COMMAND
S11
DISCARD ENTIRE STRING MOVE CURSOR TO FIRST CHARACTER
ACCEPT & FINISH COMMAND
S12
ACCEPT CURRENT CHARACTER AND STORE COMPLETED STRING AS COMPLETED ENTRY
FINISH COMMAND
S13
ACCEPT ALL CHARACTERS APART FROM CURRENT CHARACTER AS COMPLETED ENTRY
CANCEL OPERATION COMMAND
S14
DISCARD ENTIRE INPUT STRING
STOP

METHOD FOR ENTERING CHARACTERS

The present invention concerns data input and is particularly, though not exclusively, concerned with devices which have a very limited number of keys.

With the increasing miniaturisation of electronic devices it has become increasingly difficult to provide full keyboards in the available space.

A typical example of such a device would be a mobile phone. Another example is the Home-Fax (TM) manufactured by Canon Inc which can store names and associated phone numbers in a "phonebook" memory. In such a device entering text and numbers is a relatively complex operation because of the limited key-set available.

Thus many small devices without a keyboard would benefit from the ability to enter text and digits by voice. However, the replacement of key operated entry of names and numbers by a speech recognition interface raises a number of difficulties which are made worse when the devices are intended to be both small and relatively inexpensive.

Firstly the audio quality on such devices may be low. There are several reasons for this. Obviously in an inexpensive device cost is a limiting factor. Additionally on a mobile phone the background noise may be loud. For example the user could be on a train. Additionally the environment is likely to be different each time the device is used making accurate modelling of the noise characteristics difficult. Where a hands-free microphone is used the distance between the user and the microphone may limit the device's ability to reject background noise. Since the purpose of having a speech interface is to simplify a task that is carried out relatively infrequently it may not be acceptable to require a user to go through a training procedure. Thus another difficulty is that a speaker independent solution may be required. Furthermore, the device may also be used by several users again making a completely speaker independent solution preferable.

Yet another problem of inexpensive devices is that the processing capabilities available on the device may be limited by cost, battery life requirements and other factors such as its physical size. Thus as well as the complexity of the calculations that may be performed as part of the recognition algorithm being limited, the amount of working memory may also be limited as may the amount of pre-calculated data available to the program.

For the above reasons a speech interface for entering small amounts of text on a consumer device is a difficult task and recognition accuracy is likely to be low.

A concern of the present invention is to provide an efficient user interface which can make the best use of the results from a recognition engine of limited accuracy.

In accordance with the present invention there is provided a data input system comprising means for receiving an input speech signal representing a character, means for processing the received speech signal and for comparing the processed speech signal with a stored library of characters in order to allot a character from the library to the input speech signal;

means for displaying the allotted character as a candidate character, and means responsive to another input from a user of the device to store the displayed character as part of a retrievable data stream, and wherein if the user decides that the displayed character does not correspond to the initial input speech signal, a repetition of the input speech signal by the user in the absence a selection input causes the processing and display means to display another candidate character from the stored library which is different from the previously displayed character, this process being repeatable until a selection input from the user selects a display character as a correct character.

In order that the present invention may be more readily understood an embodiment thereof will now be described by way of example and with reference to the accompanying drawings in which:

Figure 1 is a perspective view of a desk-top fax system;

Figure 2 is a block diagram of a voice recognition sub-system of a device; and

Figure 3 is a flow diagram setting out the basic steps of the operation of the block diagram shown in Figure 2.

Referring now to Figure 1 of the accompanying drawings this shows a Home-Fax system having a main casing indicated at 1 and a telephone handset 2. In the present embodiment there is provided a numeric keypad 3 having push buttons representing the ten integers 0 to 9 together with * and # inputs making a total of twelve buttons. The main body also carries other inputs such as mode select switches and ON/OFF buttons. These are generally indicated at 4. The fax shown in Figure 1 is capable of storing for example 100 telephone numbers and the names associated with those numbers. In order to alleviate the problems already discussed concerning manual input with a limited keyboard the fax includes a simple microphone indicated at 5 and a liquid crystal alpha numeric display shown at 6.

The apparatus employs a speech recognition algorithm.

Having appropriately switched the machine so that it is ready to accept speech input the user says a letter, syllable or digit which is to be entered in the fax memory. Such an input will hereinafter be referred to as a character. The speech recognition interface analyses the speech input and displays a candidate character in response to the speech input. The current input is highlighted on the display 6 as representing the latest result. This differentiates the current input from any other characters which may be displayed. This can be done in any one of a number of appropriate ways. For example the candidate character can be displayed with a cursor beneath it, or as a flashing display or with an inverted colour. Of course any other method of highlighting can be used.

If the display result is incorrect the user merely repeats the original digit. The speech recognition interface then updates the display. The new result is calculated with the prior knowledge that the previously displayed result was incorrect. It is also possible in a variation to utilise the previous speech which had been input.

On the other hand, if the candidate character was correct or the repeated digit was then correctly identified the user can either say a keyword or press an appropriate key in order to indicate that the candidate character is the required character. Having selected the character, and if data entry has been completed the user may now say another keyword or again press a key indicating that data entry is over so as to store the displayed string of characters for subsequent use. Alternatively if another letter, syllable or digit is to be entered the user continues by saying it.

The above steps represent in general the normal mode of operation of the algorithm.

In addition to these fundamental features the algorithm may include a number of other features which can be used to control the process. These may include:

Firstly, a keyword or key-press to cancel the whole operation;

Secondly, a keyword or key-press to restart entering the current item from the beginning. The item could be the current word, phone number, line of text etc;

Thirdly a keyword or key-press to go back to the previous character and start entering this character again;

Fourthly a keyword or key-press to indicate that the process of entering the current character should be restarted. One reason for doing this would be that the user has mistakenly corrected the desired character;

Fifthly a keyword or key-press to go back to the previous result displayed for the current character. Again this would be done if the current character had been mistakenly replaced with a new character;

Sixthly a keyword or key-press both to accept the current character being displayed and to indicate that data entry has been completed. This procedure might be preferable than the use of two steps; and

Finally a keyword or key-press to clear the current character, leaving the system in a state where either the character can be reentered or the previous characters can be accepted as the complete entry.

With regard to the above it will be assumed that the set of keywords used can be recognised with greater accuracy than the digits, syllables or letters which are to be entered. This can be achieved because the keywords can be chosen in advance to be sufficiently distinct from the remaining vocabulary. The algorithm may also require that the keywords be recognised with a higher degree of confidence as compared to the other words in the vocabulary.

Having now given a general description of the speech recognition algorithm reference will now be made to Figure 2 of the accompanying drawings which is a block diagram of a speech recognition interface used in the embodiment of Figure 1. Of course this speech recognition interface can be used in a wide range of other devices such as mobile phones where it is required to enter alpha-numeric data with a limited number of keys.

Thus the speech recognition interface shown in Figure 2 comprises a microphone 10 the output of which is taken to a filtering and AD conversion circuit 11 where the analog signal from the microphone is filtered to remove unwanted frequencies and converted into digital format for processing in a central processor unit indicated at 12. Most of the information useful for speech recognition is contained in the frequency band between 300Hz and 4KHz. Therefore, filter portion of circuit 11 removes all frequencies outside this frequency band. Since no information which is useful for speech recognition is filtered out by the filter there is no loss of recognition performance. Further, in some environments, for example in a motor vehicle, most of the background noise is below 300Hz and the filtering can result in an effective increase in signal-to-noise ratio of approximately 10dB or more. The filtered speech signal is then converted into 16 bit digital samples by the analogue-to-digital component of circuit 11. To adhere to the Nyquist sampling criterion, the circuit samples the filtered signal at a rate of 8000 times per second.

In the central processor unit 12 the digitised speech is subjected to feature extraction by a feature extraction circuit 13. The concept of extracting features from input voice data and using the extracted features to match templates is well known and the number of features extracted and the degree of detail involved will depend on the computational power and memory space available. Basically the functions of the feature extraction circuit 13 are to extract the information required from the input speech and to reduce the amount of data that has to be processed in a recognition engine 14. There are many different types of information which can be extracted from the input signal. In this embodiment the feature extraction circuit 13 is designed to extract "formant" related information. Formants are defined as being the resonant frequencies of the vocal tract of the user, which change as the shape of the vocal tract changes.

The extracted features are stored in a buffer 15 which forms part of the interface memory 16. The memory

16 can be a single memory chip compartmentalised by an addressing regime or it can comprise a number of independent memory chips.

The memory 16 also includes a memory area 17 containing a list of active word models. This list will accordingly contain all keywords as well as the alpha numeric characters, symbols or digits which are to be recognised by the speech recognition interface. Finally the memory 16 also contains a storage area 18 housing data representing noise levels and spectral characteristics of the noise. This data is representative of background noise. The output of the three memory areas 15, 17 and 18 are all connected to a recognition engine 14.

It will be appreciated that because of its particular requirements the vocabulary of the present embodiment is relatively small. For example in the English language there are 10 digits and 26 characters. These will have to be supplemented by the chosen keywords. Japanese, for example requires nearer 110 characters and digits.

As with the feature extraction section of the speech recognition interface the operation of the recognition engine in comparing the extracted features with the list of active word models is entirely conventional.

In a preferred embodiment all previous speech input examples from buffer 15 are used to generate the new result having temporarily removed all known incorrect results from the vocabulary as already described. Thus in this preferred embodiment, when an utterance by the user is run through the speech recognition interface this generates a score for each word remaining in the vocabulary and for each previous utterance. In the simplest case the scores for each remaining in the vocabulary, from all input examples are added. The word which ends up with the best score becomes the next recognition result. More complex alternatives for combining the scores include taking the median value, discarding the words value for each word and then adding the remaining values and so on.

The results of the operation of the recognition engine 14 are supplied to the control algorithm section 19. The detailed operation of the control algorithm section 19 will be described in greater detail in the subsequent description of the flow diagram of Figure 3.

A keyboard 20 is connected to the control algorithm section and the latter is in turn connected to a display 21 and to the buffer area 15 of memory 16 and the list of active word models area 17 of memory 16. As will be appreciated the microphone, keyboard and display corresponds to the similar units as those shown on the embodiment of Figure 1.

Turning now to the flow diagram of Figure 3 this starts at step S1 with a reset step in which buffer 15 is cleared and the list of active word models in memory area 17 is reset so that all active words are present in the memory area 17.

At step S2 speech recognition is carried out in response to audio input. As already described the recognition engine 14 may make use of all previous inputs for the current character. Step S3 is a decision step where it is decided whether the output of the recognition engine is a recognised digit or a recognised keyword or an equivalent keyboard input has occurred.

If a digit has been recognised the algorithm proceeds to step S4 where the recognised digit is displayed on the display as a candidate character. Again as described this displayed character is distinguishable from the others as its correctness has not yet been confirmed. In step S5 the digit just recognised is removed from the list of active word models being used in the recognition process and in step S6 the data used for recognition of the digit is stored. The stored data could be the raw digitised audio data or the extracted features used by the recognition engine.

If the displayed candidate character is correct the user indicates this, as previously described, with a keyword or by an appropriate keystroke. If it is not the user indicates this by repeating the audio input. The recognition engine repeats the recognition procedure but of course because of step S5 cannot identify the newly input audio data with the discarded character.

Once an input digit has been correctly identified as already described the recognition procedure is then continued for the next input character via steps S2, S3, S4, S5 and S6 until the recognition engine recognises a keyword or equivalent keyboard input which does not represent a digit to be recognised. In such a case the B path of the flow diagram is followed. The first option on this path is step S7 in which the previously displayed result is displayed as the current candidate result. The purpose of this step is to ensure that a user can return to a previous character if for any reason this character has been by-passed. Step S8 follows if the user accepts a display digit. In this step the highlighting of the previously provisional digit is removed and the cursor is moved to the next character. Step S9 represents the situation if the user has requested a reset of a character by an appropriate keystroke or keyword. In this step the currently displayed output is removed. At step S10 the interface responds to a "back one character" command input via an appropriate keystroke and an appropriate keyword and acts to remove the character at the current cursor position (if present) and move the cursor on the display back one character and also remove that character for possible re-entry. Step S11 occurs in response to a restart line command and in this step the entire string displayed is discarded and the cursor is moved to the first character position of the display so that a new string can be started.

The final three steps do not involve the operation of the recognition engine with step S12 responding to an accept and finish command to accept the currently displayed character and store the completed string as a completed entry.

Step S13 is the equivalent of step S12 except that all characters apart from the current character are stored as a completed entry. Finally S14 is a response to a cancel operation command in which the entire string which had previously been input is discarded.

The system just described has a number of advantages in that successive attempts at recognising each character can use all the information from previous recognition attempts. For example the previous results can be taken out of the vocabulary and the previous speech inputs and the new input fed to the recognition engine. Even with very inaccurate recognition engines the process is guaranteed to produce the correct result eventually as incorrect results are gradually eliminated, starting with the ones which are most likely to be confused. Additionally the process of correcting an error is extremely simple in that the user has only to repeat the digit to be recognised. Finally by making the correction process part of the normal process of entering a character the user is not left with the impression that the device is failing to perform.

It will be appreciated that the overhead of having to accept each character becomes more significant when recognition is good and less significant when recognition is bad. This will make differences in performance of the recognition engine less noticeable, for example when there are differing background noise levels.

CLAIMS:

1. A data input system comprising means for receiving an input speech signal representing a character, means for processing the received speech signal and for comparing the processed speech signal with a stored library of characters in order to allot a character from the library to the input speech signal;

means for displaying the allotted character as a candidate character, and means responsive to another input from a user of the device to store the displayed character as part of a retrievable data stream, and wherein if the user decides that the displayed character does not correspond to the initial input speech signal, a repetition of the input speech signal by the user in the absence a selection input causes the processing and display means to display another candidate character from the stored library which is different from the previously displayed character, this process being repeatable until a selection input from the user selects a display character as a correct character.

2. A system according to claim 1, wherein in the absence of a selection input to a displayed character and in response to another input from the user the processing and display means are adapted to delete from the choices

available the stored library that character the display of which was not followed by the selection input.

3. A system according to claim 2, wherein once a selection input has been given in response to a displayed character, any characters which may have been barred from selection from the stored library are again made available for selection.

4.. A system according to claim 1 or 3, wherein if the user has already selected at least one previous character for storage, a candidate character is displayed in a manner which is differentiated from the or each already selected character.

5. A system according to any one of claims 1 to 4, wherein said processing means are adapted, when a character has not been selected, to utilise the speech input for the non-selected candidate characters as well as the repeated speech input to allot a new character from the library to the speech input.

6. A system according to any one of the preceding claims, wherein said library is adapted to store a plurality of alpha-numeric characters.

7. A system according to any preceding claim, wherein a candidate character is selected by the utterance of a specified keyword.

8. A system according to claim 7, wherein the confidence level for the recognition of the keyword is higher than the confidence level required for the recognition of characters for the allocation of characters.

9. A system according to any one of claims 1 to 6, wherein a candidate character can be selected by a keystroke.

10. A system according to any preceding claim, wherein a candidate character can be selected either by a keyword or a keystroke.

11. A data speech input system substantially as hereinbefore described with reference to Figures 2 and 3 of the accompanying drawings.

CHARTER MARK

PATENTS • DESIGNS
The Patent Office
COPYRIGHT • TRADE MARKS

INVESTOR IN PEOPLE

10

Application No:	GB 0017887.1	**Examiner:**	John Donaldson
Claims searched:	1 to 11	**Date of search:**	3 August 2001

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.S): G4R(RHA)

Int Cl (Ed.7): G10L 15/00, 15/22

Other: Online:WPI, EPODOC, JAPIO

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
X	GB 2098773 A	(ASULAB), see page 3, line 49 to page 4, line 11, page 4, lines 40 to 57	1, 5, 9, 10
X	US 5781887	(JUANG), see column 3, lines 22 to 52	1, 5, 6

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.